Randomized Quasi-Monte Carlo for MCMC

Radu Craiu [1] Christiane Lemieux [2]

[1]Department of Statistics, Toronto

[2]Department of Statistics, Waterloo

Third Workshop on Monte Carlo Methods
Harvard, May 2007

Outline

Local Search and Optimization Algorithms (LSOA)

- Search locally a region of the state space to sample from.
- Adapt locally to improve the efficiency of the algorithm.
- Local Search and Optimization Algorithms (LSOA):
  Multiple-Try Metropolis (Liu , Liang and Wong, JASA 2000),
  Hit-and-Run (Chen and Schmeiser, JCGS 1993),
  Adaptive Direction Sampling (Gilks, Roberts and George, Stat. 1994),
  Delayed rejection (Green and Mira, Biomka, 2001),
  Directional Metropolis (Eidsvik and Tjelmeland , Statist. & Comput. 2006).

Idea: Produce a more structured search using stratified sampling.
Today's Topic: Discussion of RQMC implemented within LSOA.

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw K independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw K independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x,y)$ is a symmetric function.

(i) Draw K independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

(iii) Accept y with probability $\min\left\{1, \frac{\sum_{i=1}^K w(y_i|x_t)}{\sum_{i=1}^K w(x_i^*|y)}\right\}$.

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw K independent trial proposals $y_1, \ldots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \ldots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

(iii) Accept y with probability $\min\left\{1, \frac{\sum_{i=1}^K w(y_i|x_t)}{\sum_{i=1}^K w(x_i^*|y)}\right\}$.

- Do we better explore the sample space with K proposals ?

Original MTM (Liu, Liang and Wong, JASA 2000)

- Suppose T is a proposal distribution such that $T(x|y) > 0 \Leftrightarrow T(y|x) > 0$ and $\lambda(x, y)$ is a symmetric function.

(i) Draw K independent trial proposals $y_1, \dots, y_K$ from $T(\cdot|x_t)$. Sample one with $p_i \propto w(y_i|x_t) = \pi(y_i)T(x_t|y_i)\lambda(x_t, y_i)$.

(ii) Generate $x_1^*, \dots, x_{k-1}^* \sim T(\cdot|y)$ and put $x_k^* = x_t$.

(iii) Accept y with probability $\min\left\{1, \frac{\sum_{i=1}^K w(y_i|x_t)}{\sum_{i=1}^K w(x_i^*|y)}\right\}$.

- Do we better explore the sample space with K proposals ?
- As long as we take advantage of the flexibility provided by the MTM.

Multi-distributed MTM (MD-MTM)

- The proposals do not have to be identically distributed: $y_j \sim T_j(\cdot|x)$ for $1 \leq j \leq k$.

Multi-distributed MTM (MD-MTM)

- The proposals do not have to be identically distributed: $y_j \sim T_j(\cdot|x)$ for $1 \le j \le k$.

If $y = y_{j_0}$ is selected than put $x^*_{j_0} = x_t$ and sample $x^*_j \sim T_j(\cdot|y)$.

- May accomodate different sampling regimes.

Multi-distributed MTM (MD-MTM)

- The proposals do not have to be identically distributed: $y_j \sim T_j(\cdot|x)$ for $1 \le j \le k$.

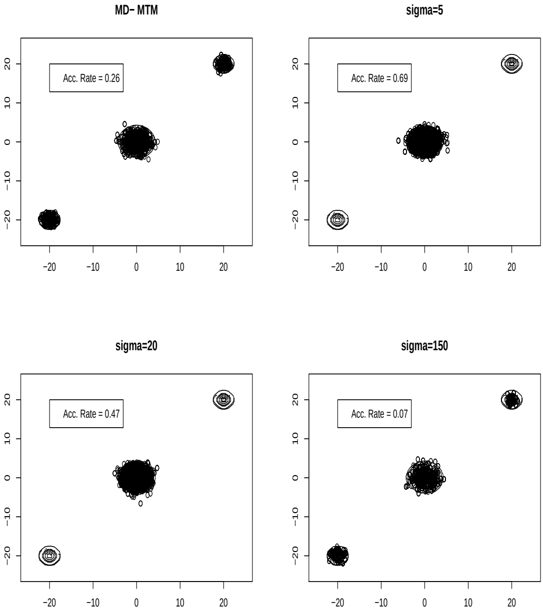
If $y = y_{j_0}$ is selected than put $x^*_{j_0} = x_t$ and sample $x^*_j \sim T_j(\cdot|y)$.

- May accomodate different sampling regimes.

$$\pi(x) = 0.3N((-20,-20)^T, \Sigma_1) + 0.4N((0,0)^T, \Sigma_2) + 0.3N((20,20)^T, \Sigma_3),$$

where $\Sigma_1 = \mathrm{diag}(0.5, 0.5)$, $\Sigma_2 = \mathrm{diag}(2, 2)$ and $\Sigma_3 = \mathrm{diag}(0.4, 0.4)$.

- Use MD-MTM for a random walk Metropolis with Gaussian proposal and variances $\sigma \in \{5, 50, 100, 150, 200\}$.
- Compare with MTM for a random walk Metropolis constructed with only one of the σ's.


MD- MTM
Acc. Rate = 0.26
sigma=5
Acc. Rate = 0.69
sigma=20
Acc. Rate = 0.47
sigma=150
Acc. Rate = 0.07
−20
−10
0
10
20

Multiple-Correlated-Try Metropolis (MCTM)

- The proposals do not have to be independent.

Multiple-Correlated-Try Metropolis (MCTM)

- The proposals do not have to be independent.

MCTM(I)

Suppose we sample K trial proposals $y_1, \ldots, y_K$ from $\tilde{T}(y_1, \ldots, y_K | x_t)$ where

$$\int \tilde{T}(y_1, \ldots, y_K | x_t) dy_{[-i]} = T(y_i | x_t), \ \ \forall 1 \le i \le K.$$

- The proposals do not have to be independent.

MCTM(I)

Suppose we sample K trial proposals $y_1, \ldots, y_K$ from $\tilde{T}(y_1, \ldots, y_K | x_t)$ where

$$\int \tilde{T}(y_1, \ldots, y_K | x_t) dy_{[-i]} = T(y_i | x_t), \quad \forall 1 \leq i \leq K.$$

- The algorithm proceeds as in the independent case with one exception.

Multiple-Correlated-Try Metropolis (MCTM)

- The proposals do not have to be independent.

MCTM(I)

Suppose we sample K trial proposals $y_1, \ldots, y_K$ from $\tilde{T}(y_1, \ldots, y_K|x_t)$ where

$$\int \tilde{T}(y_1, \ldots, y_K|x_t) dy_{[-i]} = T(y_i|x_t), \quad \forall 1 \leq i \leq K.$$

- The algorithm proceeds as in the independent case with one exception.

MCTM (II)

Draw $(x_1^*, \ldots, x_{K-1}^*)$ variates from the conditional transition kernel $\tilde{T}(x_1, \ldots, x_{K-1}|y, x_K = x_t)$ and let $x_K^* = x_t$.

- The proposals do not have to be independent.

MCTM(I)

Suppose we sample K trial proposals $y_1, \ldots, y_K$ from $\tilde{T}(y_1, \ldots, y_K | x_t)$ where

$$\int \tilde{T}(y_1, \ldots, y_K | x_t) dy_{[-i]} = T(y_i | x_t), \quad \forall 1 \leq i \leq K.$$

- The algorithm proceeds as in the independent case with one exception.

MCTM (II)

Draw $(x_1^*, \ldots, x_{K-1}^*)$ variates from the conditional transition kernel $\tilde{T}(x_1, \ldots, x_{K-1} | y, x_K = x_t)$ and let $x_K^* = x_t$.

- We have freedom in choosing $\tilde{T}$ as long as we can perform the conditional sampling step.

Antithetic or Stratified Sampling?

- $Y_j = (Y_{j1}, \ldots, Y_{jd}) \in R^d$ for $1 \leq j \leq K$.

Antithetic or Stratified Sampling?

- $Y_j = (Y_{j1}, \ldots, Y_{jd}) \in R^d$ for $1 \leq j \leq K$.

Say $K = 2$, $\mathrm{Var}(Y) = \Sigma = diag(\sigma_1^2, \ldots, \sigma_d^2)$, $\mathrm{Cov}(Y_{1h}, Y_{2h}) = \sigma_h^2 \rho_h$. Then

$$E\left[||Y_1 - Y_2||^2\right] = E\left[\sum_{h=1}^{d} (Y_{1h} - Y_{2h})^2\right] = \sum_{h=1}^{d} \sigma_h^2 (1 - \rho_h).$$

Antithetic or Stratified Sampling?

- $Y_j = (Y_{j1}, \ldots, Y_{jd}) \in R^d$ for $1 \leq j \leq K$.

Say $K = 2$, $\text{Var}(Y) = \Sigma = diag(\sigma_1^2, \ldots, \sigma_d^2)$, $\text{Cov}(Y_{1h}, Y_{2h}) = \sigma_h^2 \rho_h$. Then

$$E\left[||Y_1 - Y_2||^2\right] = E\left[\sum_{h=1}^{d} (Y_{1h} - Y_{2h})^2\right] = \sum_{h=1}^{d} \sigma_h^2 (1 - \rho_h).$$

- One possible choice is $\rho_h = -1/(K-1)$ (covariance matrix needs to be positive definite!).
- The optimal choice of ρ_h depends on σ_h, $1 \leq h \leq d$.

Antithetic or Stratified Sampling?

- $Y_j = (Y_{j1}, \dots, Y_{jd}) \in R^d$ for $1 \le j \le K$.

Say $K = 2$, $\mathrm{Var}(Y) = \Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)$, $\mathrm{Cov}(Y_{1h}, Y_{2h}) = \sigma_h^2 \rho_h$. Then

$$E\left[||Y_1 - Y_2||^2\right] = E\left[\sum_{h=1}^{d} (Y_{1h} - Y_{2h})^2\right] = \sum_{h=1}^{d} \sigma_h^2 (1 - \rho_h).$$

- One possible choice is $\rho_h = -1/(K-1)$ (covariance matrix needs to be positive definite!).
- The optimal choice of ρ_h depends on σ_h, $1 \le h \le d$.
- Conditional sampling needed for MCTM is straightforward in the case of Gaussian proposals.

Antithetic or Stratified Sampling?

- $Y_j = (Y_{j1}, \dots, Y_{jd}) \in R^d$ for $1 \le j \le K$.

Say $K = 2$, $\mathrm{Var}(Y) = \Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)$, $\mathrm{Cov}(Y_{1h}, Y_{2h}) = \sigma_h^2 \rho_h$. Then

$$E\left[||Y_1 - Y_2||^2\right] = E\left[\sum_{h=1}^{d} (Y_{1h} - Y_{2h})^2\right] = \sum_{h=1}^{d} \sigma_h^2 (1 - \rho_h).$$

- One possible choice is $\rho_h = -1/(K-1)$ (covariance matrix needs to be positive definite!).
- The optimal choice of ρ_h depends on σ_h, $1 \le h \le d$.
- Conditional sampling needed for MCTM is straightforward in the case of Gaussian proposals.
- If $Y_j = \psi(X_t, U_j)$ with ψ increasing in U_j then we need to use U's which are negatively associated (NA - Craiu and Meng, AnnStat. 2005).

Quasi-Monte Carlo (QMC)

- QMC is a de-randomized MC; the samples are constructed deterministically.
- QMC: highly uniform and stratified sampling in the unit hypercube.
- Interest: high-uniformity, a.k.a. low discrepancy.

Quasi-Monte Carlo (QMC)

- QMC is a de-randomized MC; the samples are constructed deterministically.
- QMC: highly uniform and stratified sampling in the unit hypercube.
- Interest: high-uniformity, a.k.a. low discrepancy.

For $v \in (0,1)^d$, $B(v) = \{u \in (0,1)^d : 0 < u_j \leq v_j,\ 1 \leq j \leq d\}$.

- QMC is a de-randomized MC; the samples are constructed deterministically.
- QMC: highly uniform and stratified sampling in the unit hypercube.
- Interest: high-uniformity, a.k.a. low discrepancy.

For $v \in (0,1)^d$, $B(v) = \{u \in (0,1)^d : 0 < u_j \leq v_j,\ 1 \leq j \leq d\}$.

For a point set S_n define $\alpha(S_n, v) = \text{card}\{S_n \cap B(v)\}$.

Quasi-Monte Carlo (QMC)

- QMC is a de-randomized MC; the samples are constructed deterministically.
- QMC: highly uniform and stratified sampling in the unit hypercube.
- Interest: high-uniformity, a.k.a. low discrepancy.

For $v \in (0,1)^d$, $B(v) = \{u \in (0,1)^d : 0 < u_j \leq v_j,\ 1 \leq j \leq d\}$.

For a point set S_n define $\alpha(S_n, v) = \text{card}\{S_n \cap B(v)\}$.

Star-discrepancy

$$D_n^*(S_n) = \sup_{v \in (0,1)^d} |v_1 \ldots v_d - \alpha(S_n, v)/n|.$$

Randomized QMC (RQMC)

- Efficiency can be connected to discrepancy.

Randomized QMC (RQMC)

- Efficiency can be connected to discrepancy.

Koksma-Hlawka Theorem

$$\left| \mu - \frac{1}{n} \sum_{u \in S_n} f(u) \right| \leq D_n^* V(f).$$

- Both $V(f)$ and D^* are difficult to compute in general.

Randomized QMC (RQMC)

- Efficiency can be connected to discrepancy.

Koksma-Hlawka Theorem

$$\left| \mu - \frac{1}{n} \sum_{u \in S_n} f(u) \right| \le D_n^* V(f).$$

- Both $V(f)$ and D^* are difficult to compute in general.
- Randomized versions of QMC (RQMC) algorithms are used to establish error estimates via Monte Carlo simulation. (Art Owen, Stanford; Christiane Lemieux, Waterloo; Pierre L'Ecuyer, Montréal).
- Randomization is performed so that it preserves the low-discrepancy of the set point.
- Makes QMC set points suitable for MC as marginals are uniform.
- RQMC and MCMC (Owen and Tribble, PNAS '05; Lemieux and Sidorsky, Math. & Comp. Model., 2006).

Latin Hypercube Sampling (LHS)

- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \le i \le K$, $1 \le j \le d$:

Latin Hypercube Sampling (LHS)

- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \le i \le K$, $1 \le j \le d$:
  i) $V_{ij} \sim \text{Uniform}(0,1)$ are i.i.d

Latin Hypercube Sampling (LHS)

- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \le i \le K$, $1 \le j \le d$:
  i) $V_{ij} \sim \text{Uniform}(0,1)$ are i.i.d
  ii) $\{\tau_j, \ \ 1 \le j \le d\}$ are independent permutations of $\{0, \ldots, K-1\}$

Latin Hypercube Sampling (LHS)

- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \leq i \leq K$, $1 \leq j \leq d$:
  i) $V_{ij} \sim \text{Uniform}(0,1)$ are i.i.d
  ii) $\{\tau_j,\ \ 1 \leq j \leq d\}$ are independent permutations of $\{0, \ldots, K-1\}$

$$U_{ij} = \frac{\tau_j(i-1) + V_{ij}}{K}.$$

Latin Hypercube Sampling (LHS)

- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \leq i \leq K$, $1 \leq j \leq d$:
  i) $V_{ij} \sim \text{Uniform}(0,1)$ are i.i.d
  ii) $\{\tau_j,\ \ 1 \leq j \leq d\}$ are independent permutations of $\{0,\dots,K-1\}$

$$U_{ij} = \frac{\tau_j(i-1) + V_{ij}}{K}.$$

- $U^{(i)}$ is the ith row of the matrix U.

Latin Hypercube Sampling (LHS)

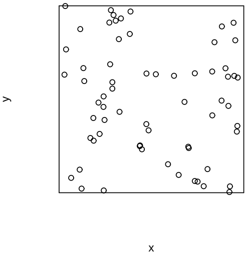
- Simultaneous stratification of the univariate marginals if $U \sim \text{Uniform}(0,1)^d$.
- **Construction:** For $1 \le i \le K$, $1 \le j \le d$:
  i) $V_{ij} \sim \text{Uniform}(0,1)$ are i.i.d
  ii) $\{\tau_j,\ \ 1 \le j \le d\}$ are independent permutations of $\{0, \dots, K-1\}$

$$U_{ij} = \frac{\tau_j(i-1) + V_{ij}}{K}.$$

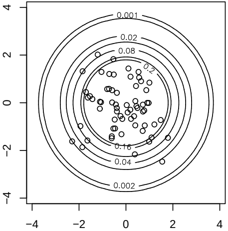
- $U^{(i)}$ is the ith row of the matrix U.
- Given $U^{(i)} = (U_{i1}, \dots, U_{id}) \in (0,1)^d$, $\ 1 \le i \le K$ we can use:
  a) For $1 \le j \le d$ fixed: $\{U_{1j}, \dots, U_{(Kj)}\}$ as a set of NA Uniform$(0,1)$ r.v.'s.
  b) $(U^{(1)}, ..., U^{(K)})$ as a stratified sample of $(0,1)^d$.

LHS and Gaussian sampling: $K = 64$, $d = 2$

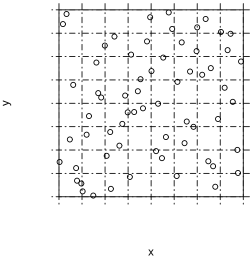
Independent



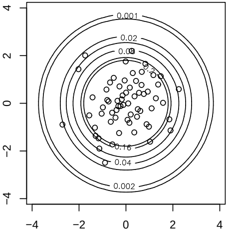
Bivariate normal



LHS



Bivariate normal



LHS in MTM

LHS in MTM

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \leq i \leq K-1$ (no permutation is needed!).

LHS in MTM

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \leq i \leq K-1$ (no permutation is needed!).
- If $Y = Y_{i_0}$ is selected:

LHS in MTM

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \le i \le K-1$ (no permutation is needed!).
- If $Y = Y_{i_0}$ is selected:

Intuitively, if Y has been generated using a uniform from the interval , $(\frac{i_0}{K}, \frac{i_0+1}{K})$, then we want to sample $x_2^*, \ldots, x_k^*$ using all the $K-1$ uniforms which ARE NOT IN $(\frac{i_0}{K}, \frac{i_0+1}{K})$.

LHS in MTM

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \le i \le K-1$ (no permutation is needed!).
- If $Y = Y_{i_0}$ is selected:

Intuitively, if Y has been generated using a uniform from the interval , $(\frac{i_0}{K}, \frac{i_0+1}{K})$, then we want to sample $x_2^*, \ldots, x_k^*$ using all the $K-1$ uniforms which ARE NOT IN $(\frac{i_0}{K}, \frac{i_0+1}{K})$.

i) Put $j_0 = [KF_y(X_t)]$ (where $[\cdot]$ is integer part) and construct a random permutation τ such that $\tau(j_0) = i_0$.

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \leq i \leq K-1$ (no permutation is needed!).
- If $Y = Y_{i_0}$ is selected:

Intuitively, if Y has been generated using a uniform from the interval , $(\frac{i_0}{K}, \frac{i_0+1}{K})$, then we want to sample $x_2^*, \ldots, x_k^*$ using all the $K-1$ uniforms which ARE NOT IN $(\frac{i_0}{K}, \frac{i_0+1}{K})$.

i) Put $j_0 = [KF_y(X_t)]$ (where $[\cdot]$ is integer part) and construct a random permutation τ such that $\tau(j_0) = i_0$.
ii) Put $x_{j_0}^* = X_t$, and for $i \neq j_0$ $U_i^* = \frac{\tau(i)+W_i}{K}$ and $x_i^* = F_y^{-1}(U_i^*)$, with $W_i \sim \text{Uniform}(0,1)$ independent from the V's.

- $V_i \sim U(0,1)$, $U_i = \frac{i+V_i}{K}$, $Y_i = F_{X_t}^{-1}(U_i)$, $0 \le i \le K-1$ (no permutation is needed!).
- If $Y = Y_{i_0}$ is selected:

Intuitively, if Y has been generated using a uniform from the interval , $(\frac{i_0}{K}, \frac{i_0+1}{K})$, then we want to sample $x_2^*, \dots, x_k^*$ using all the $K-1$ uniforms which ARE NOT IN $(\frac{i_0}{K}, \frac{i_0+1}{K})$.

i) Put $j_0 = [KF_y(X_t)]$ (where $[\cdot]$ is integer part) and construct a random permutation τ such that $\tau(j_0) = i_0$.
ii) Put $x_{j_0}^* = X_t$, and for $i \neq j_0$ $U_i^* = \frac{\tau(i)+W_i}{K}$ and $x_i^* = F_y^{-1}(U_i^*)$, with $W_i \sim$ Uniform$(0,1)$ independent from the V's.

- One can get away without permuting at all!!

Korobov Rules

- K points in $[0,1]^d$ are generated using a generator $a \in \{1, \ldots, K-1\}$.

Korobov Rules

- K points in $[0,1]^d$ are generated using a generator $a \in \{1, \ldots, K-1\}$.

$$S_K = \left\{ \frac{i}{K}(1, a, a^2 \mod K, \ldots, a^{d-1} \mod K) \mod 1 \right\}_{0 \le i \le K-1}.$$

Korobov Rules

- K points in $[0,1]^d$ are generated using a generator $a \in \{1,\ldots,K-1\}$.

$$S_K = \{\tfrac{i}{K}(1, a, a^2 \mod K, \ldots, a^{d-1} \mod K) \mod 1\}_{0 \le i \le K-1}.$$

- Bad choices: a and K are not relatively prime or $a = 1$.
- The point set is *fully projection regular*, i.e. any lower-dimensional projection contains K distinct points.

Korobov Rules

- K points in $[0,1]^d$ are generated using a generator $a \in \{1, \ldots, K-1\}$.

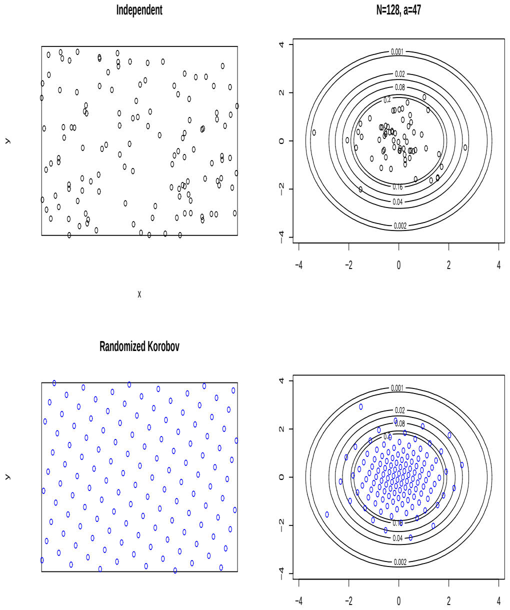
$$S_K = \left\{ \frac{i}{K}(1, a, a^2 \mod K, \ldots, a^{d-1} \mod K) \mod 1 \right\}_{0 \le i \le K-1}.$$

- Bad choices: a and K are not relatively prime or $a = 1$.
- The point set is *fully projection regular*, i.e. any lower-dimensional projection contains K distinct points.

To randomize add a vector $U_K \sim \text{Uniform}(0,1)^d \mod 1$ i.e.

$$S_K^R = S_K + U_K \mod 1.$$

Example: Korobov rule and Gaussian sampling



Going beyond uniformity

Going beyond uniformity

- Sometimes we want to "favor" regions in the hypercube but uniformity is lost once we transform the QMC point set.

Going beyond uniformity

- Sometimes we want to "favor" regions in the hypercube but uniformity is lost once we transform the QMC point set.
- $g : [0,1] \rightarrow [0,1]$ is a bijection and $\tilde{\mathbf{U}} = g(\mathbf{U}) = (g(U_1), \ldots, g(U_d))$.
- $Y \sim T(\cdot|X) \Leftrightarrow Y = \psi(X, \mathbf{U})$ so if $Y = \psi(X, \tilde{\mathbf{U}})$ then $Y \sim \tilde{T}(\cdot|X)$.

Going beyond uniformity

- Sometimes we want to "favor" regions in the hypercube but uniformity is lost once we transform the QMC point set.
- $g : [0,1] \to [0,1]$ is a bijection and $\tilde{\mathbf{U}} = g(\mathbf{U}) = (g(U_1), \ldots, g(U_d))$.
- $Y \sim T(\cdot|X) \Leftrightarrow Y = \psi(X, \mathbf{U})$ so if $Y = \psi(X, \tilde{\mathbf{U}})$ then $Y \sim \tilde{T}(\cdot|X)$.

If $\lambda(X,Y) = T(Y|X)/\tilde{T}(Y|X)$ is symmetric, i.e. $\lambda(X,Y) = \lambda(Y,X)$, then MCTM can be implemented. without computing $\tilde{T}$.

Going beyond uniformity

- Sometimes we want to "favor" regions in the hypercube but uniformity is lost once we transform the QMC point set.
- $g : [0,1] \to [0,1]$ is a bijection and $\tilde{\mathbf{U}} = g(\mathbf{U}) = (g(U_1), \ldots, g(U_d))$.
- $Y \sim T(\cdot|X) \Leftrightarrow Y = \psi(X, \mathbf{U})$ so if $Y = \psi(X, \tilde{\mathbf{U}})$ then $Y \sim \tilde{T}(\cdot|X)$.

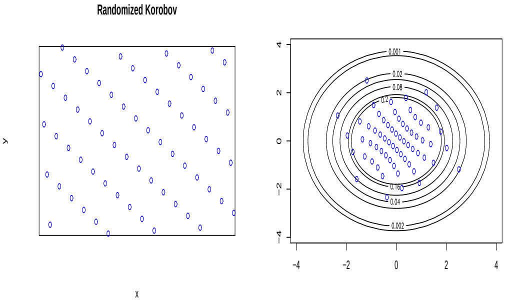
If $\lambda(X,Y) = T(Y|X)/\tilde{T}(Y|X)$ is symmetric, i.e. $\lambda(X,Y) = \lambda(Y,X)$, then MCTM can be implemented. without computing $\tilde{T}$.

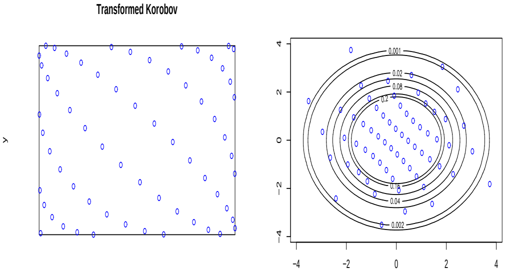
- **Example**:

$$g(u) = \frac{\sin[\pi(u-0.5)] + 1}{2}.$$

i) g is bijective.
ii) $g(x) = x,\ \forall x \in \{0, 1/2, 1\}$, $g(u) + g(1-u) = 1,\ \forall u \in (0,1)$.

Example


Randomized Korobov
y
x
0.001
0.02
0.08
0.2
0.16
0.04
0.002
-4
-2
0
2
4


Transformed Korobov
y
0.001
0.02
0.08
0.2
0.16
0.04
0.002
-4
-2
0
2
4

Gaussian Proposals

Suppose $T(y|x)$ is a multivariate Gaussian with no correlation.

$$T(y|x) = \prod_{j=1}^{d} \frac{e^{-(y_j - x_j)^2/2\sigma_j^2}}{\sqrt{2\pi\sigma_j^2}}$$

Then

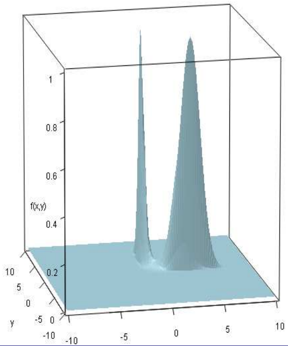
$$\tilde{T}(y|x) = \prod_{j=1}^{d} \frac{e^{-(y_j - x_j)^2/2\sigma_j^2}}{\sqrt{2\pi\sigma_j^2}} \frac{1}{\pi\sqrt{\Phi(\frac{y_j - x_j}{\sigma_j})\left[1 - \Phi(\frac{y_j - x_j}{\sigma_j})\right]}}.$$

$$\lambda(y, x) = \prod_{j=1}^{d} \frac{1}{\pi\sqrt{\left[1 - \Phi(\frac{x_j - y_j}{\sigma_j})\right]\Phi(\frac{x_j - y_j}{\sigma_j})}} = \lambda(x, y).$$

Bimodal density

Use MCTM combined with random-ray Monte Carlo to sample from (Gelman and Meng, Am. Stat. '91)

$$\pi(x_1, x_2) \propto \exp\{-(9x_1^2x_2^2 + x_1^2 + x_2^2 - 8x_1 - 8x_2)/2\}$$



Efficiency Comparison

- At each iteration a random direction v is generated;
- Along direction v we sample proposals $y_1, \dots, y_k$ around the current state x_t using $y_i = x_t + r_i$ with $r_i \sim \text{Uniform}(-\sigma, \sigma)$.

Table: *Values of the MSE reduction factor* $R = \frac{MSE_{anti}}{MSE_{ind}}$.

$\sigma \backslash k$	3	4	5	6
3	0.35	0.53	0.64	0.81
4	0.31	0.42	0.58	0.76
5	0.29	0.40	0.49	0.62

Lupus Data

- Lupus Data (van Dyk and Meng, JCGS 2001): 55 patients, 2 covariates; logistic model.
- The posterior density is proportional to

$$\pi(\beta|x,y) \propto \prod_{j=0}^{2} \frac{e^{-0.5\beta_j/100^2}}{100\sqrt{2\pi}} \prod_{i=1}^{55} \left[\frac{\exp(X_i^T\beta)}{1+\exp(X_i^T\beta)}\right]^{y_i} \left[\frac{1}{1+\exp(X_i^T\beta)}\right]^{1-y_i}.$$

- Sample using MCTM with antithetically coupled proposals or stratified proposals via a randomized Korobov set shifted using the transformation g.

Efficiency comparison for Lupus Data

Table: *Values of R for β_1 and $p_{25} = 1_{\{\beta_1 > 25\}}$ in the logit example.*

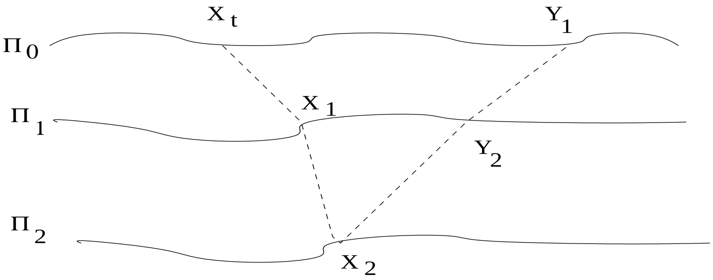
	Antithetic			QMC		
$k \backslash \sigma$	2	3	4	2	3	4
3	0.92/0.92	0.90/0.86	0.99/0.95	-	-	-
4	0.94/0.87	0.88/0.88	0.91/0.89	-	-	-
5	0.98/0.96	0.81/0.81	0.89/0.86	-	-	-
6	0.91/0.86	0.86/0.78	0.95/0.92	-	-	-
8	0.81/0.70	0.75/0.69	0.83/0.80	0.69/0.72	0.61/0.60	0.59/0.56
16	0.87/0.81	0.97/0.94	0.91/0.88	0.81/0.81	0.82/0.84	0.76/0.75

Table: Computation Times $k = 8$

Sample Size	MTM	MCTM (A)	MCTM (S)
10^5	34	36	32

Multi-Annealed Metropolis (MAM)

- Inspired by Neal's (Stat. & Comp., 1996) tempered transition method.
- Uses a series of distributions $\{\pi_t\}_{t \in \{1=T_0, T_1, \ldots, T_M\}}$, constructed between the distribution of interest, π, and π_{T_M}.
- Relies on a "path" that crosses all the intermediate distributions.
- If $1 = T_0 < T_1 < T_2 = T$ and Q_{T_i} is a proposal distribution for $\pi_{T_i} \propto \pi^{1/T_i}$.
- Suppose that $Q_{T_i}(X|Y) = Q_{T_i}(Y|X)$ (e.g., random walk Metropolis).

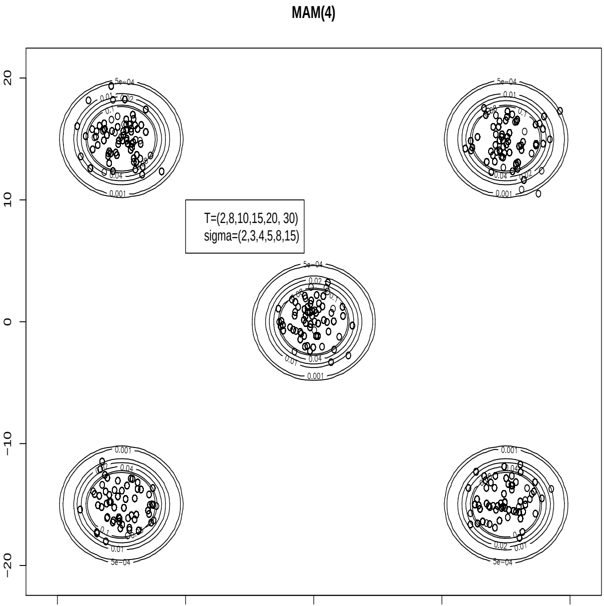


Given the chain is in state X:

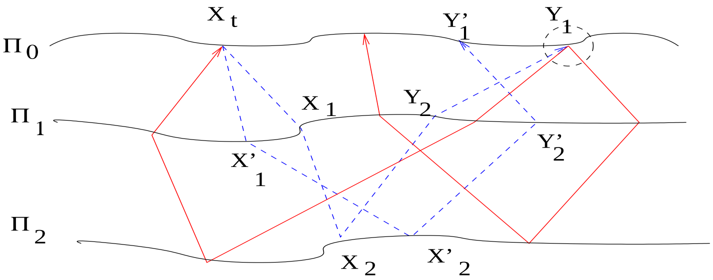
i) Sample $X_1 \sim Q_1(\cdot|X)$, $X_2 \sim Q_2(\cdot|X_1)$, $Y_2 \sim Q_2(\cdot|X_2)$, $Y_1 \sim Q_1(\cdot|Y_2)$.

ii) Accept Y_1 with probability $\min\left\{1, \frac{\pi_0(Y_1)}{\pi_0(X)} \cdot \frac{\pi_2(X_1)\pi_1(Y_2)}{\pi_1(X_1)\pi_2(Y_2)}\right\}$.

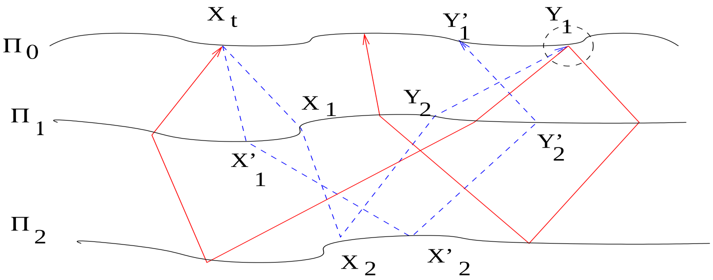
Example


MAM(4)
T=(2,8,10,15,20, 30)
sigma=(2,3,4,5,8,15)
20
10
0
−10
−20
5e−04
0.01
0.02
0.1
0.04
0.001

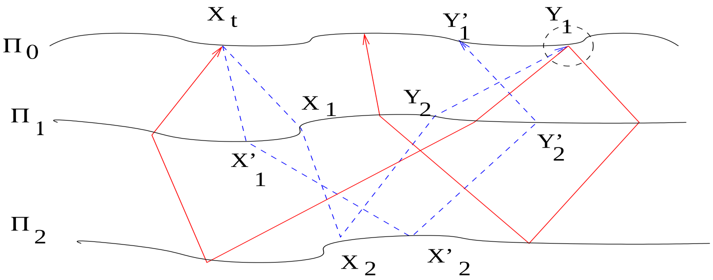
Multiple-try extension of the MAM



Multiple-try extension of the MAM



MTM can be implemented by generating multiple paths and compensate using the same number of reverse paths. We have freedom in choosing the "weight of a path".



MTM can be implemented by generating multiple paths and compensate using the same number of reverse paths. We have freedom in choosing the "weight of a path".

MAM

Pros : increases the acceptance rate >10 fold, does not require running the parallel chains all the time;
Cons: Requires symmetric proposals.

Negative Association for Multiple-Try MAM

- Suppose that $X_i \sim Q_i(\cdot|X_{i-1}) \Leftrightarrow X_i = \psi_i(X_{i-1}, U_i)$ with $U_i \sim \text{Uniform}[0,1]^d$ and ψ_i is monotone in each component of U_i.
- $Y_1 = \psi_1(\psi_2(\psi_2(\psi_1(X, U_1), U_2), V_2), V_1)$ and $Y_1' = \psi_1(\psi_2(\psi_2(\psi_1(X, U_1'), U_2'), V_2'), V_1')$.
- The paths do not need to be generated independently.
- If (U_i, U_i') and (V_i, V_i') are negatively associated then $\text{Corr}(Y_1, Y_1') < 0$ so the average distance between the paths is larger than under independence.